

A consensus approach for estimating the predictive accuracy of dynamic models in biology

Alejandro F Villaverde¹, Sophia Bongard², Klaus Mauch², Dirk Müller², Eva Balsacanto¹, Joachim Schmid², Julio R Banga¹

¹ Bioprocess Engineering Group, IIM-CSIC. Eduardo Cabello 6, 36208 Vigo, Spain
{afvillaverde, ebalsa, julio}@iim.csic.es

² Insilico Biotechnology AG, Meitnerstraße 8, 70563 Stuttgart, Germany
{sophia.bongard, klaus.mauch, dirk.mueller, joachim.schmid}@insilico-biotechnology.com

Abstract

Mathematical models that predict the complex dynamic behaviour of cellular networks are fundamental in systems biology, and provide an important basis for biomedical and biotechnological applications. However, obtaining reliable predictions from large-scale dynamic models is commonly a challenging task due to lack of identifiability. The present work addresses this challenge by presenting a methodology for obtaining high-confidence predictions from dynamic models using time-series data. First, to preserve the complex behaviour of the network while reducing the number of estimated parameters, model parameters are combined in sets of meta-parameters, which are obtained from correlations between biochemical reaction rates and between concentrations of the chemical species. Next, an ensemble of models with different parameterizations is constructed and calibrated. Finally, the ensemble is used for assessing the reliability of model predictions by defining a measure of convergence of model outputs (consensus) that is used as an indicator of confidence. We report results of computational tests carried out on a metabolic model of Chinese Hamster Ovary (CHO) cells, which are used for recombinant protein production. Using noisy simulated data, we find that the aggregated ensemble predictions are on average more accurate than the predictions of individual ensemble models. Furthermore, ensemble predictions with high consensus are statistically more accurate than ensemble predictions with large variance. The procedure provides quantitative estimates of the confidence in model predictions and enables the analysis of sufficiently complex networks as required for practical applications.

Keywords. Systems biology, Metabolic engineering, Cell line engineering, Dynamic modeling, Ensemble modeling, Consensus.

1 Introduction

Mathematical modeling is a fundamental task in systems and computational biology [1], with important applications in biomedicine [2–5]. Among other features, models allow monitoring the state of unmeasured variables and making predictions about system behaviour for a larger number and broader variety of conditions than can be efficiently tested in experiments [6]. The construction and calibration of models of large, complex dynamic systems is a particularly challenging task. Uncertainties appear at different stages of the process, limiting the confidence in the resulting predictions [7,8]. Shortage of experimental data can easily lead to poor identifiability. As an example, consider a well-known result from nonlinear systems theory which states that, to identify a model described by differential equations containing r parameters, $2r+1$ experimental measures may be enough [9]. This result assumes exact, noise-free measurements; however, in practice there will always be errors in the data, hence the $2r+1$ figure represents a lower bound. When the number of parameters is larger than what can be actually determined from data, the calibration procedure can sometimes—when allowed by the model structure—yield a perfect fit between model predictions and measurements. However, there is a danger of overfitting in this situation, i.e., the model is being trained to fit in detail the noise contained in the data instead of actually learning the system dynamics. This problem entails the risk that model predictions will be wrong for altered experimental conditions.

The problem of dealing with uncertainty in cellular network modeling was reviewed in [10]. In that review, the use of ensembles—sets of models with different structures and/or parameter values—was considered as a powerful and generally applicable approach for reducing prediction errors. However, it was also acknowledged that the concept has not sufficiently matured yet. Indeed, ensemble modeling approaches have been recently applied to a variety of problems, ranging from climate prediction [11] to impact of vaccines [12]. An early example of the use of an ensemble approach in biological models was presented in [13], which was limited to ensembles of topologies of Boolean networks. Tran *et al* [14] extended the approach to the dynamic case, building an ensemble of metabolic models that reached the same steady state and applying it to the central carbon metabolism of *E. coli*. A related application was presented in [15]. For a review of metabolic ensemble modeling see [16].

The use of the consensus as an indication of the reliability of the predictions was explored by Bever [17], who computed time-dependent probability distributions of protein concentrations in artificial gene regulatory networks and introduced the concept of consensus sensitivity, finding that consensus among ensemble models was a good indicator of high-confidence predictions. Recently, further steps were taken with the introduction of the concept of “core prediction”: a property that must be fulfilled if the model structure is to explain the data, even if the individual parameters are not accurately identified [18].

The present paper deals with the problem of evaluating and, if possible, increasing the confidence in the predictions made by kinetic metabolic models. It is assumed that the model structure—the topology of the metabolic network—is known. Actually, this assumption is not a requirement of the proposed methodology, which may be applied

to ensembles of models with different topologies. However, in the present work the uncertainty in the predictions is due only to uncertainty in the parameter values. To overcome uncertainty, an ensemble of models with different parameterizations is built. As a preceding step to improve identifiability and to reduce overfitting, the initial model parameters are grouped into modules of meta-parameters, which are used during calibration. Then a measure of consensus among model outcomes is introduced, which is used to quantify the confidence in the predicted metabolite concentrations. A schematic depiction of the methodology is shown in Figure 1.

We note that, while a consensus approach was proposed in [17], it used different measures than the ones we introduce here, and it was applied to toy models consisting of 3 or 4 genes. The present methodology includes entirely new features such as the use of meta-parameters, and it is tested on a medium-size network including 34 metabolites. We also remark that, unlike the approach presented in [18], we do not intend to characterize the model's core predictions, but instead to give estimates of the confidence in the predictions. Finally, we note that a preliminary version of this work [19] was presented at the PACBB'14 conference. This new version has been extensively rewritten, including new figures and results, which have been calculated with a new dissensus measure that enables a more sensitive discrimination of larger and smaller prediction errors.

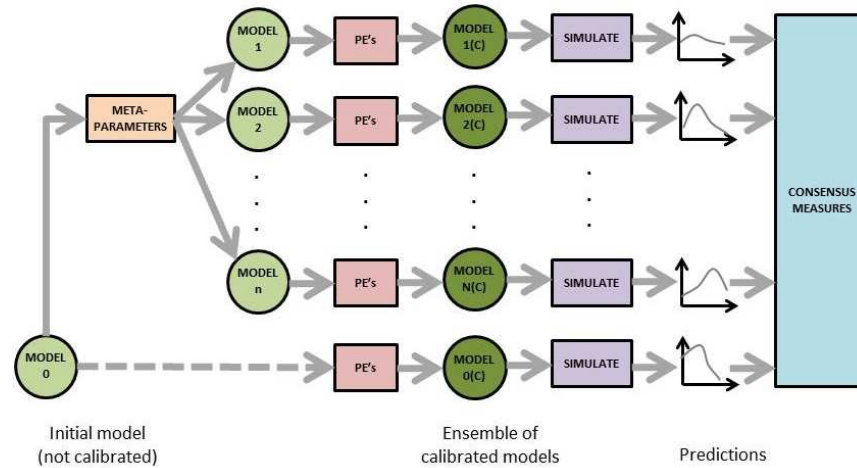


Fig. 1 - Diagram of the methodology. From the initial model ("Model 0") different meta-parameter sets are obtained, leading to a set of models with different parameterizations ("Model 1", "Model 2", ..., "Model n"). Parameter estimation (PE) is then performed on these models, leading to an ensemble of calibrated models ("Model 1(C)", ..., "Model n(C)"). Additionally, the original model can also be calibrated without using the meta-parameter approach (dashed grey line leading to "Model 0 (C)"). The resulting ensemble of models is simulated, obtaining an ensemble of predictions. From these predictions, consensus measures are calculated and conclusions are obtained.

2 Methods

2.1. Meta-parameter approach

The methodology aims at adapting the kinetics of interrelated reaction pathways. Highly correlated trajectories of simulated concentrations and reaction rates point at functional dynamic relations, which can be adjusted by the parameters that correspond to the correlated time courses of concentrations and fluxes. We will refer to these sets of parameters as meta-parameters and use them for improving identifiability and reducing the risk of overfitting.

Let us consider an ordinary differential equation (ODE) model with rate kinetics which follow the description in [20], where a rate of an enzyme i is defined by

$$r_i = r_i^{max} \cdot f_i(c_j, p_i^j) = r_i^{max} \cdot \left(1 + \sum_{j=1}^{npars} p_i^j \ln\left(\frac{c_j}{c_j^0}\right)\right) \quad (1)$$

with r_i being a product of the maximal rate (r_i^{max}) and a kinetic rate expression (f_i) which is a function of the metabolite concentrations (c_j) and the parameters (p_i^j) that quantify the contribution of metabolite j in the rate i . The function f_i is following here linlog kinetics [21], but it can be any generic kinetic rate equation where parameters p_i^j are associated to specific metabolite concentrations c_j , such as K_M values in Michaelis-Menten type kinetic equations.

The model is simulated with a set of initial parameter values $p_i^{j0} = (p_i^{10}, \dots, p_i^{npars0})$, obtaining time courses of concentrations and rates. The Pearson Correlation Coefficients (PCC) are then calculated between simulated concentration time courses for all balanced species c_j , as well as between all simulated rates r_i . This yields two correlation arrays, C_r and C_c :

$$C_r = \begin{pmatrix} PCC_{r1,r1} & \cdots & PCC_{r1,rm} \\ \vdots & \ddots & \vdots \\ PCC_{rm,r1} & \cdots & PCC_{rm,rm} \end{pmatrix}, C_c = \begin{pmatrix} PCC_{c1,c1} & \cdots & PCC_{c1,cn} \\ \vdots & \ddots & \vdots \\ PCC_{cn,c1} & \cdots & PCC_{cn,cn} \end{pmatrix} \quad (2)$$

Meta-parameters are defined by pairing metabolites and reactions. First, a set of metabolites is selected, $c_{j\text{ sel}}$, which is of particular interest, e.g. because their time courses have been measured. Each metabolite of interest is paired with the reaction rate with the highest contribution to the metabolite balance, $r_{i\text{ sel}}$. In most cases the flux with the highest contribution corresponds to the formation rate of the metabolite. Rates which are highly correlated ($PCC > 0.8$) with the selected rate $r_{i\text{ sel}}$ are then included in a vector $r_{i\text{ corr}}$. Similarly, the metabolites which are present in the kinetic equations of the correlating rates ($r_{i\text{ corr}}$) and which are highly correlated with $c_{j\text{ sel}}$ are included in a vector, $c_{j\text{ corr}}^i$. Thus for each concentration-rate pair ($c_{j\text{ sel}}, r_{i\text{ sel}}$), two vectors $c_{j\text{ corr}}^i$ and $r_{i\text{ corr}}$ are obtained.

A meta-parameter vector $MP_{i\ sel,j\ sel}$ consists of all the parameters $p_{i\ corr}^{j\ corr}$ which are related to the metabolite concentrations $c_{j\ corr}^{i\ corr}$ in the correlating rates $r_{i\ corr}$. Each meta-parameter vector $MP_{i\ sel,j\ sel}$ is split into two vectors, one for substrate parameters $MP_{i\ sel,j\ sel}^{subs}$ and one for product parameters $MP_{i\ sel,j\ sel}^{prod}$:

$$MP_{i\ sel,j\ sel}^{prod} = (p_{i\ corr}^{j\ prod\ corr}), MP_{i\ sel,j\ sel}^{subs} = (p_{i\ corr}^{j\ subs\ corr}) \quad (3)$$

If identical meta-parameters are identified from correlation analysis, just one parameter vector is kept to avoid co-linearity.

For each meta-parameter, an optimization parameter $k_{i\ sel,j\ sel}$ is defined, which multiplies every parameter in the set. This reduces the size of the optimization problem: instead of estimating all the parameters $p_{i\ corr}^{j\ corr}$ included in a meta-parameter separately, they are kept at their nominal values and only the $k_{i\ sel,j\ sel}$ need to be estimated, that is, $MP_{i\ sel,j\ sel}^{prod\ optimal} = MP_{i\ sel,j\ sel}^{prod\ nominal} \times k_{i\ sel,j\ sel}^{prod\ optimal}$ for product parameters, and $MP_{i\ sel,j\ sel}^{subs\ optimal} = MP_{i\ sel,j\ sel}^{subs\ nominal} \times k_{i\ sel,j\ sel}^{subs\ optimal}$ for substrate parameters.

Multiple meta-parameters can be grouped into modules. A module is a set of meta-parameters that is designed to manipulate the dynamics of one phenotypic property. With a different choice of metabolites of interest, different meta-parameter sets for the same model structure are obtained. Thus, the property of interest determines the choice of meta-parameters.

In the present work the meta-parameter sets have been determined using the modelling and simulation software Insilico Discovery™ (Insilico Biotechnology AG, Stuttgart, Germany). The routine has a low computational cost: for the model under examination, calculations were carried out in less than three minutes.

For better understanding, in the remaining of this subsection we explain the generation and grouping of a meta-parameter for a particular example. Let us assume that we want to obtain a module of meta-parameters related to the energy metabolism, in which we know that e.g. glucose plays a key role. We therefore choose glucose (c_{glc}) as $c_{j\ sel}$. The glucose uptake rate ($r_{glc\ uptake}$) has the highest contribution to the node balance, and it is therefore chosen as $r_{i\ sel}$. Then, we generate the concentration and rate correlation arrays by calculating the Pearson Correlation Coefficients from simulated concentration and rate time curves. We obtain the vector $r_{glc\ uptake\ corr}$ for rates which correlate strongly (i.e., PCC > 0.8) with $r_{glc\ uptake}$. Concentrations which are (i) included in the kinetic expressions of the rates in $r_{glc\ uptake\ corr}$ and (ii) correlate strongly with c_{glc} , are included in the vector $c_{glc\ corr}$. One of the rates in the rate vector $r_{glc\ uptake\ corr}$ is the aspartate-glutamate carrier reaction (r_{agc}). The kinetic equation for this reaction is shown in equation (4) below. The kinetic rate expression (r_{agc}) includes the concentrations of four metabolites: two substrates, cytosolic glutamate (c_{glu_c}) and mitochondrial aspartate (c_{asp_m}), and two products, mitochondrial glutamate (c_{glu_m}) and cytosolic aspartate (c_{asp_c}):

$$r_{agc} = r_{agc}^{max} \left(1 + p_{agc}^{subsglu_c} \ln \left(\frac{c_{glu_c}}{c_{glu_c}^0} \right) + p_{agc}^{subsp_m} \ln \left(\frac{c_{asp_m}}{c_{asp_m}^0} \right) + p_{agc}^{prod\ glum} \ln \left(\frac{c_{glu_m}}{c_{glu_m}^0} \right) + p_{agc}^{prod\ aspc} \ln \left(\frac{c_{asp_c}}{c_{asp_c}^0} \right) \right) \quad (4)$$

From the four metabolites present in the equation, we find one (cytosolic glutamate) whose concentration time course, c_{glu_c} , is strongly correlated with the time course of the glucose concentration, c_{glc} ; hence we include c_{glu_c} in the vector $c_{glc\ corr}^{glc\ uptake\ corr}$. The parameter corresponding to both r_{agc} and c_{glu_c} is $p_{agc}^{subs\ glu_c}$, which is therefore included in the meta-parameter vector $MP_{glc\ uptake, glc}^{subs}$ (since cytosolic glutamate is a substrate in the reaction of the aspartate-glutamate carrier). By repeating this procedure for all the rates present in the vector $r_{glc\ uptake\ corr}$, we obtain the meta-parameters $MP_{glc\ uptake, glc}^{subs}$ and $MP_{glc\ uptake, glc}^{prod}$. The set of parameters included in those meta-parameters, together with the parameters in other concentration—rate pairs (e.g. c_{ATP_c} - r_{atpase}) that are key metabolites for the energy metabolism, form the module “energy metabolism”.

As already mentioned, the procedure is not only valid for linlog kinetics, but also for any generic kinetic rate equation. To demonstrate the validity for an alternative rate equation, we reformulate equation (4) using irreversible Michaelis-Menten kinetics. Using the notation $K_M = p_i^j$ for the Michaelis-Menten constant, we have the following expression:

$$r_{agc} = r_{agc}^{max} \left(\left(\frac{c_{aspm}}{c_{aspm} + p_{agc}^{subs\ aspm}} \right) \cdot \left(\frac{c_{glu_c}}{c_{glu_c} + p_{agc}^{subs\ glu_c}} \right) \right) \quad (5)$$

Following the same reasoning as before, since the concentration c_{glu_c} is highly correlated with the concentration of our metabolite of interest, c_{glc} , the corresponding parameter $p_{agc}^{subs\ glu_c}$ is included in the meta-parameter $MP_{glc\ uptake, glc}^{subs}$. This is equivalent to the linlog case.

2.2. Calibrating the models in the ensemble

The parameter values are estimated by calibrating the models in the ensemble with the enhanced scatter search parameter estimation method [22]. It is a state-of-the-art metaheuristic that has shown competitive performance with large-scale multimodal problems [23]. A parallel version of the method is available [24]; however, the computational effort associated with the optimization problem tackled in the present work is not as demanding as to require parallel techniques. Hence we decided to apply the sequential version of the method that is included in the AMIGO toolbox [25]. Recent applications of parallel optimization metaheuristics in bioinformatics can be found e.g. in [26].

The enhanced scatter search metaheuristic (eSS) can be regarded as a hybrid technique which combines global and local search, the latter in order to accelerate convergence to nearby optima. In this application, the local method chosen is the Matlab routine *fmincon*, available in the Optimization Toolbox.

When there is lack of identifiability, different sets of parameter values can fit the data equally well. Because eSS is a stochastic technique, several optimization runs with different random initializations result in different solutions, each of which is a member of the ensemble. Models in the ensemble share the same structure but have different parameterizations. Each meta-parameter set leads to a model with a different parameter set for optimization purposes. In turn, each of these models can give rise to several models by changing the parameter values. The resulting ensemble includes (i) models calibrated with the meta-parameter approach, and (ii) models that were calibrated with the original parameter set.

2.3. Consensus approach

After creating an ensemble of models with different meta-parameter set and/or different parameter values, it is used for estimating the reliability of the predictions. The behaviour of each model in a different experimental condition is simulated. Convergence of the model outputs (consensus) is then taken as an indicator of the confidence in the prediction (see section 3 for justification).

Let us introduce the following notation: let n_p be the number of data points (that is, of time instants corresponding to the measurements), n_m the number of models in the ensemble, and n_s the number of dynamic states (concentrations) for which a prediction is made. The prediction of the dynamic state j at time i made by model k is A_{ijk} . The average prediction of all the models for every state and time instant is the $n_p \times n_s$ array:

$$\overline{A_{ij}} = \frac{1}{n_m} \sum_{k=1}^{n_m} A_{ijk} \quad (6)$$

The dissensus d_{ijk} of every model k with respect to the average prediction for a concentration j at every time instant i can be encapsulated as a $n_p \times n_s \times n_m$ array,

$$d_{ijk} = \frac{(A_{ijk} - \overline{A_{ij}})^2}{\max_k(A_{ijk}) - \min_k(A_{ijk})} \quad (7)$$

The mean dissensus of every model for every state along the time horizon is

$$d_{jk} = \frac{\sum_{i=1}^{n_p} d_{ijk}}{n_p} \quad (8)$$

And the ensemble dissensus is the mean dissensus among all the models for a particular state along the time horizon,

$$d_j = \frac{\sum_{k=1}^{n_m} d_{jk}}{n_m} \quad (9)$$

From which we define the ensemble consensus for each state as:

$$x_j = \max(d_j) - d_j \quad (10)$$

To compare the consensus elicited by different states in the same model it is useful to use normalized metrics which have values between 0 and 1. With this aim we define a scaled ensemble consensus for each state as:

$$sx_j = \frac{\max(d_j) - d_j}{\max(d_j)} \quad (11)$$

In this way, if we obtained e.g. $sx_1 = 0.8051$, $sx_2 = 0.9874$, and $sx_3 = 0.9903$, we would note that the ensemble consensus for state variable 2 (which typically represents the concentration of a chemical species) is clearly higher than for state variable 1, and therefore the methodology would suggest that the predictions about the concentrations of species 2 are more reliable than for species 1. Furthermore, since the differences between sx_2 and sx_3 are relatively small, we could also conclude that the predictions about the concentrations of state variables 2 and 3 are similarly reliable.

Finally, we can calculate an average global consensus for all the states as

$$x = \frac{\sum_{j=1}^{n_s} x_j}{n_s} \quad (12)$$

Defining other consensus metrics is straightforward, by choosing other measures of the distance between model predictions. For example, in equation (7) we could use the difference in absolute values rather than the squared difference, i.e.

$$d_{ijk} = \frac{|A_{ijk} - \bar{A}_{ij}|}{\max_k(A_{ijk}) - \min_k(A_{ijk})} \quad (13)$$

The expression described in equation (13) was chosen in a previous work [19]. However, comparisons between both metrics have shown that using the expression in equation (7) enables a more sensitive discrimination of larger and smaller prediction errors, and is hence the one used in the present work.

We quantify the error in the prediction of model k with respect to the true value of a concentration j at every time instant i as:

$$e_{ijk} = \frac{(A_{ijk} - A_{ij}^*)^2}{\max_k(A_{ijk})} \quad (14)$$

where A_{ij}^* is the true value of the concentration. The error of every model for every state along the time horizon is

$$e_{jk} = \frac{\sum_{i=1}^{n_p} e_{ijk}}{n_p} \quad (15)$$

And the average prediction error for a state is the average error of all the models,

$$e_j = \frac{\sum_{k=1}^{n_m} d_{jk}}{n_m} \quad (16)$$

To normalize the prediction error between 0 and 1 we define a scaled error as:

$$se_j = \frac{e_j}{\max(e_j)} \quad (17)$$

3 Results and Discussion

3.1. Case study: Chinese Hamster Ovary cells model

The methodology was applied to a metabolic model of Chinese Hamster Ovary cells (CHO), which are used for recombinant protein production in fermentation processes. We simulated a batch process with resting cells and a time horizon of 300 hours. The fermenter medium contains glucose as the main carbon source, and leucine and methionine as representative amino acids that are taken up. Lactate is modeled as key by-product of the fermentation process. A synthesized protein serves as main product of the fermentation process. Glucose, lactate, product protein, leucine, and methionine are assumed to be measured in the fermenter. Aspartate, malate, pyruvate, oxaloacetate, ATP, and ADP are assumed to be available as intracellular metabolite measurements.

The model comprises 34 metabolites (whose concentrations are the state variables, numbered in Table 1) and 32 reactions, and includes protein product formation, Embden-Meyerhof-Parnas pathway (EMP), TCA cycle, a reduced amino acid metabolism, lactate production, and the electron transport chain. The metabolic network is shown on Fig. 2 and the detailed reactions are listed in the supplementary information (Table 7). The kinetic ODE model comprises 117 parameters in total. The kinetic expressions of the rates are of the form of equation (1). The values of r_i^{max} that appear in the equation can be taken from a stationary reference flux distribution, which we obtained by applying Flux Balance Analysis (FBA). The objective function for FBA is defined as:

$$f_{obj}^{FBA} = \max(mr) \quad (18)$$

where mr is the maintenance rate, which is defined as surplus of all ATP producing fluxes over all ATP demanding metabolic processes in the cell. Uptake and production rate constraints were chosen to be close to measured stationary fluxes in [27]. The remaining parameters that need to be identified are: the K_m values in the Michaelis-Menten kinetics, and the elasticities in the linlog kinetics. The model is available in different formats (SBML, Matlab, C, COPASI, AMIGO) as part of the BioPreDyn-Bench suite [28], which can be downloaded from <http://www.iim.csic.es/~gingproc/biopredynbench/>

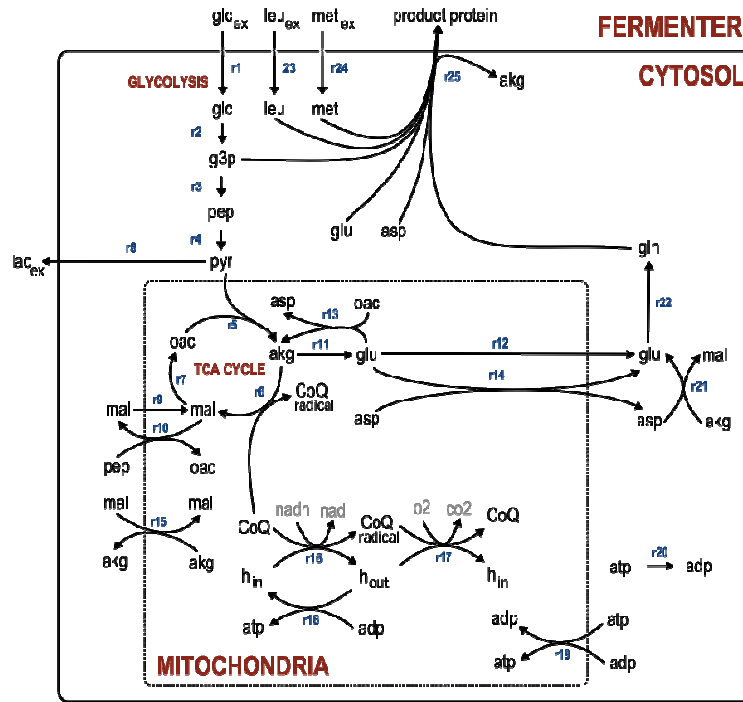


Fig. 2 - Metabolic network of the CHO cell model.

The model was subdivided into five different physiological modules, for which associated meta-parameter modules were generated (more details are given in Tables 2—6 in the Supplementary Information):

- Central metabolism module: MPs from concentration and rate pairs related to the EMP and TCA cycle, describing the dynamics of glucose, leucine, methionine, and 2-oxoglutarate.
- Fermenter module: describing the dynamics of glucose, lactate, methionine, leucine, and product protein.
- Energetics module: TCA cycle and energy-related metabolites, describing the dynamics of glucose, pyruvate, methionine, leucine, and ATP.
- Uptake and energetics module: substrate-associated meta-parameters describing the dynamics of glucose, methionine, leucine, ATP, product protein, and lactate.
- Uptakes module: describing the dynamics of glucose, methionine, and leucine.

Table 1. List of state variables (metabolite concentrations) in the CHO model. Commonly measured metabolites are marked with (*).

Nr	Variable name	Nr	Variable name
1	Glucose (fermenter) (*)	18	H_out (mitochondria)
2	Lactate (fermenter) (*)	19	CoQ (mitochondria)
3	Leucine (fermenter) (*)	20	H_in (mitochondria)
4	Methionine (fermenter) (*)	21	Pyruvate (cytosol) (*)
5	Product protein (fermenter)(*)	22	Phosphoenolpyruvate (cytosol)
6	Glutamate (mitochondria)	23	NADH (cytosol)
7	NAD (mitochondria)	24	Glycerate3-phosphate (cytosol)
8	2-Oxoglutarate (mitochondria)	25	NAD (cytosol)
9	NADH (mitochondria)	26	Glucose (cytosol)
10	Glutamine (cytosol)	27	Malate (cytosol) (*)
11	ADP (cytosol) (*)	28	2-Oxoglutarate (cytosol)
12	Glutamate (cytosol)	29	Aspartate (cytosol) (*)
13	ATP (cytosol) (*)	30	ATP (mitochondria) (*)
14	Aspartate (mitochondria)	31	Orthophosphate (mitochondria)
15	Oxaloacetate (mitochondria) (*)	32	ADP (mitochondria) (*)
16	Malate (mitochondria)	33	Leucine (cytosol)
17	CoQH_radical (mitochondria)	34	Methionine (cytosol)

Meta-parameters in these modules were estimated either for single modules or for module combinations during the parameter estimation procedure. Synthetic measurement data was generated using a reference model and was used in lieu of experimental data. Care was taken to ensure resemblance to realistic experimental conditions. For each of the 13 metabolites that are typically measured, 12 sampling times were assumed and Gaussian noise was added with a standard deviation of 20% while non-negativity for all synthetically generated concentrations was ensured.

3.2. Calibrating the CHO cells model

Using the eSS algorithm for parameter estimation included in AMIGO, we carried out 10 optimizations of the full model without meta-parameters (117 parameters) starting from different initial random guesses. The lower and upper bounds for the parameters were 1/5 and 5 times the nominal values. All of the optimizations succeeded in ob-

1
2
3
4
5
6
7
8
9
10 taining a good fit to the data, yielding 10 solutions that differed significantly; this
11 non-uniqueness suggests lack of practical identifiability.

12 Following this initial calibration with the original parameter set, we re-
13 parameterized the model using the meta-parameter approach. Every meta-parameter is
14 defined as a function of several of the original parameters, for which a default value
15 has to be set. This default value may simply be chosen as a random initial guess with-
16 in the parameter bounds. Alternatively, the result of a calibration with the original
17 parameter set can be used, if it is available. In our case, since we had already per-
18 formed 10 optimizations, we chose to use their results for setting the default values.
19 Since 10 different possible solutions were found in the first set of calibrations, we
20 averaged the 10 parameter vector values obtained and used the result as the initial
21 guess. Realizing that this simple average biases the guess towards large values, we
22 also considered a solution with smaller parameter values as second initial guess. Thus
23 two different guesses for the vector of 117 parameters were used, $pg1$ and $pg2$.

24 Next we carried out parameter estimations with each of the 5 meta-parameter mod-
25 ules defined in section 3.1. As before, the optimizations started from 10 different
26 initial points for each module; that is, the initial guesses were 10 parameter vectors
27 whose values were chosen randomly within the parameter bounds. They were carried
28 out twice: with $pg1$ and $pg2$, resulting in a total of $2 \times 10 \times 5 = 100$ optimizations. Re-
29 ducing the number of optimized parameters (from the original 117 parameters to be-
30 tween 5 and 12 meta-parameters, depending on the module) improved the identifiabil-
31 ity of the model, and the 100 optimizations yielded only 22 different solutions (that is,
32 the fraction of different solutions obtained in the optimizations was reduced from
33 100% to 22%). In particular, the “uptake and energetics” module returned just a single
34 solution with $pg1$ and another one with $pg2$. On the other hand, we obtained 4 and 6
35 different solutions, respectively, for the fermenter module. The number of solutions
36 returned by the remaining modules was between the two aforementioned results.

37 An ensemble of models was created with these 22 parameterizations plus the 10 re-
38 sulting from the optimization of the original model, that is, a total of 32 models. All of
39 these models provided a near-perfect fit to the pseudo-experimental data used for
40 calibration, as shown in Fig. 3A. The fact that the models calibrated with the meta-
41 parameter modules (which had between 5 and 12 free parameters to be optimized)
42 managed to fit the data equally well as the models calibrated with the full original set
43 (117 parameters) supports the idea that the models calibrated with the original pa-
44 rameter set had a clear risk of overfitting. The decrease in the number of optimal solu-
45 tions found with the meta-parameter approach suggests that its use reduced this risk.
46 The reduction in the number of estimated parameters also led to a decrease of the
47 computation times of the optimization procedure by up to two orders of magnitude.
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

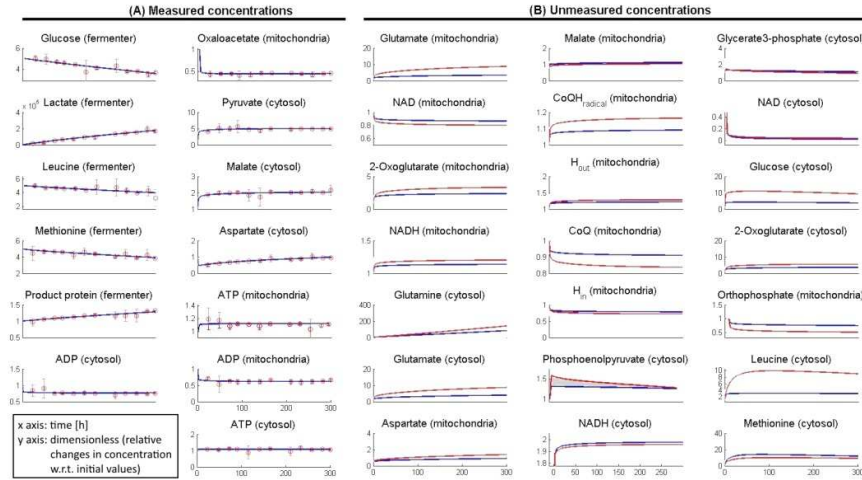


Fig. 3 – (A): Time-courses of the 13 measured state variables (metabolite concentration changes). The plots show: pseudo-experimental data (red circles with error bars); individual predictions of the 32 fitted models (grey lines); and their average—i.e., ensemble—predictions (blue lines). The pseudo-experimental data corresponding to these 13 states were used in the parameter estimation procedure.

(B): True model output with nominal parameters (red lines), individual predictions of the 32 fitted models (grey lines), and their average—i.e., ensemble—predictions (blue lines), for the 21 unmeasured states (metabolite concentrations).

3.3. Predictions of unmeasured variables

Next we evaluated the ability of the calibrated models and of the ensemble to reproduce the concentrations of the remaining 21 metabolites, which are typically not measured in practice. The scaled consensus in the predictions among models in the ensemble was calculated for each of the measured and unmeasured metabolites as in equation (11). In general, the measured metabolites elicited more consensus than the unmeasured ones as expected. Panel B of Fig. 3 shows the predictions of every model in the ensemble (in grey) together with the real output (in red) and the ensemble prediction (in blue), for the 21 unmeasured states. Visual inspection of this figure shows that consensus is a good indicator of the confidence in the predictions, a result further confirmed by the numerical results. For example, predictions of the time course of NADH (cytosol) elicit a very high consensus ($s_{x_{23}} = 0.999992$), which suggests that its consensus prediction is very close to the real output. This is indeed the case: in Fig. 3B, the ensemble prediction of this variable (blue line) is indistinguishable from its true time course (red line). Conversely, variable number 10 (concentration of glutamine in cytosol) has the lowest consensus: due to scaling, its value is actually zero,

$sx_{10} = 0$, and the resulting prediction is far from the real output. Therefore, if we were thinking of measuring additional metabolites, we should concentrate our efforts on those metabolites with the larger dissensus, such as glutamine, and we could safely avoid additional measurements and rely on the ensemble prediction for metabolites such as NADH.

Indeed, consensus and prediction error were found to be significantly anti-correlated, with a correlation coefficient of $r = -0.7719$ and p-value = $9 \cdot 10^{-8}$. Fig. 4(A) plots the scaled consensus sx_j and scaled prediction errors se_j for the 13 measured variables. Since the models were fitted to these data, the prediction errors are very low and the consensus is very high for all the variables. Fig. 4(B) shows the corresponding values for the 21 unmeasured variables.

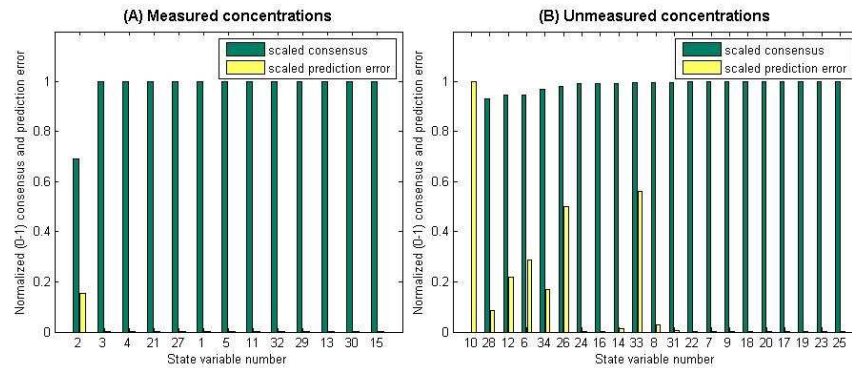


Fig. 4 – Bar plots of the scaled consensus sx_j and scaled prediction errors se_j for the 13 measured variables (panel A), and for the 21 unmeasured variables (panel B).

It can be noticed that, for the 10 unmeasured variables that elicit higher consensus (numbers 31, 22, 7, 9, 18, 20, 17, 19, 23, and 25), the prediction errors are very low, which is not always the case for the other variables. Thus, ranking the predictions according to their consensus is a useful way of deciding which metabolites should be measured and which can be left unmeasured because we have enough confidence in the predictions.

3.4. Predictions under different experimental conditions

Assessing the confidence in predictions of the system's behaviour under different conditions than those used for calibration represents another application of the consensus approach. In this case, often no experimental data are available. To test the validity of the approach in this type of scenario, we generated a new set of experimental conditions by randomly changing the initial concentrations of all the metabolites. Then we simulated the model's behaviour with nominal parameter values, generating a new artificial dataset, and compared it with the corresponding output of the ensemble models; results are shown on Fig. 5. Fig. 6 plots the scaled consensus sx_j and

scaled prediction errors se_j for the 34 variables. Again, consensus and prediction error were found to be anti-correlated, with a correlation coefficient of $r = -0.9998$ and p-value $= 5 \cdot 10^{-56}$. Finally, Fig. 7 represents in a different way the results shown in Figs. 4 and 6.

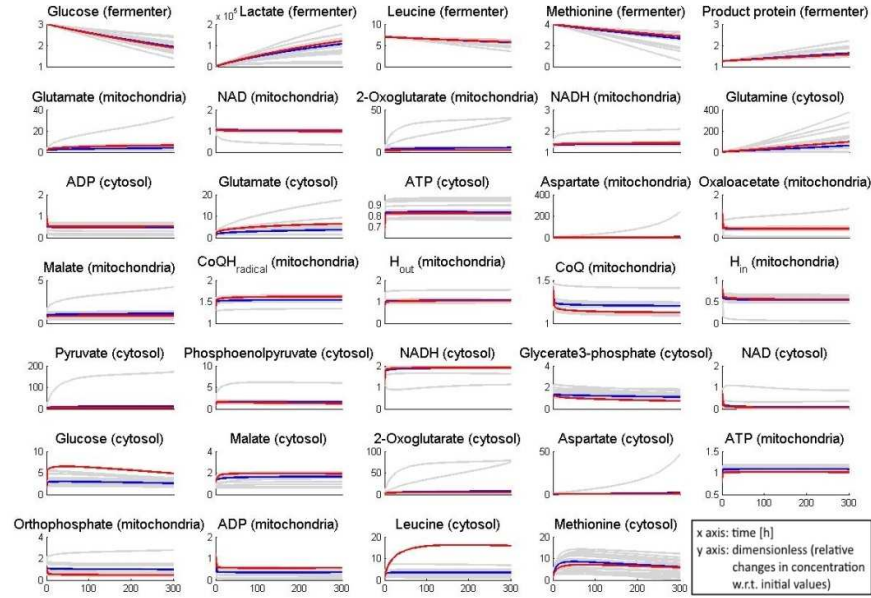


Fig. 5 - True model output with nominal parameters (red lines), individual predictions of the 32 fitted models (grey lines) and their average—i.e., ensemble—predictions (blue lines) for the 34 state variables (metabolite concentration changes) in the CHO model, under different conditions than those used in the model calibration.

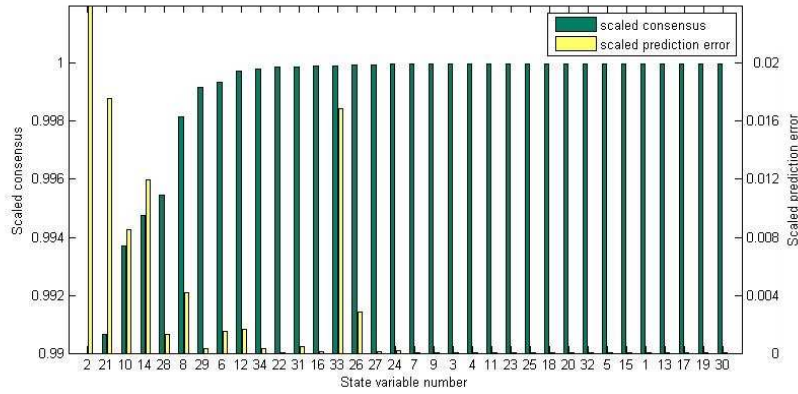


Fig. 6 - Bar plots of the scaled consensus sx_j and scaled prediction errors se_j for the 34 state variables (metabolite concentrations) in the CHO model, under different conditions than those used for calibration. Note that the scaled consensus (range 0.99—1) and scaled prediction error (range 0—0.02) refer to different axes (consensus on the left, error on the right). Note also that the scaled error lies outside the plot range: $se_2 = 1$.

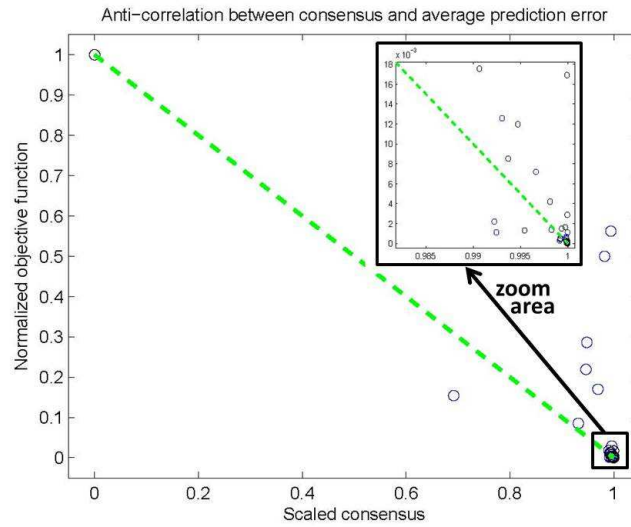


Fig. 7 – Scatter plot of scaled consensus sx_j and scaled prediction errors se_j for the 34 state variables (metabolite concentrations) in the CHO model. This figure includes the data used in Fig. 4 (blue circles) and in Fig. 6 (black circles). The dashed green line corresponds to an ideal correlation.

4 Discussion

Before proceeding to the conclusions, we discuss in this section some aspects of the procedure and the results that deserve further comment.

We begin by remarking that an advantage of using an ensemble approach is that, as suggested by the anti-correlation between sx_j and se_j , the ensemble average can generally recapitulate the true system behaviour. This is shown in figures 3 and 5, where it can be noticed that the average (blue lines) and the true output (red lines) are generally close to each other, while some individual predictions (grey lines) are sometimes very far from reality.

This is not always true, though: in some cases (i.e. for some concentrations) the ensemble average fails to recapitulate the true dynamics. However, what the aforementioned anti-correlation implies is that—to a certain extent—we can expect to know when we should trust the ensemble predictions and when we should not. Thus, from Fig. 4 we conclude that the predictions for the variables with the highest consensus (that is, 19, 23, 25...) are highly reliable; and when we look at the corresponding plots in Fig. 3 we see that this is indeed the case. Conversely, since variables such as 10, 28, and 12 elicit low consensus, we expect their predictions to be of less quality than the others, which is indeed the case if we look again at Fig. 3. To sum up, the introduction of a consensus metric complements the ensemble approach by providing indications of the confidence in the ensemble prediction.

We remark that the ensemble prediction is, on average, better than the individual predictions of the models in the ensemble. This is shown in Fig. 8, which compares the prediction errors of the ensemble average for all the states with the average error of each of the calibrated models. Furthermore, Fig. S1 (see supplementary file) aggregates the prediction errors of each model for all states, thus comparing the predictions of individual models with the ensemble prediction globally (instead of doing it separately for each state), showing that in both scenarios the ensemble prediction is among the top performers. To sum up, from the results it can be concluded that, since one cannot know a priori the quality of a particular model's prediction, the best option is always to use the ensemble prediction instead of the prediction of an individual model.

Another aspect that should be noted is that the consensus can vary between scenarios. For example, variable 21, which is pyruvate (cytosol), elicits high consensus in the first scenario (Fig. 4), while in the second scenario (Fig. 6) its consensus is very low, and as a consequence the ensemble average has a large prediction error. This difference in consensus can be easily explained: in the first scenario pyruvate (cytosol) was one of the measured variables, and hence its data were used to calibrate the models; hence all of the models managed to fit those data well—the output of all the models was nearly identical for that variable—producing a high consensus. In the second scenario, however, there were no measured data, so in principle there was no reason to expect a higher consensus for any particular variable.

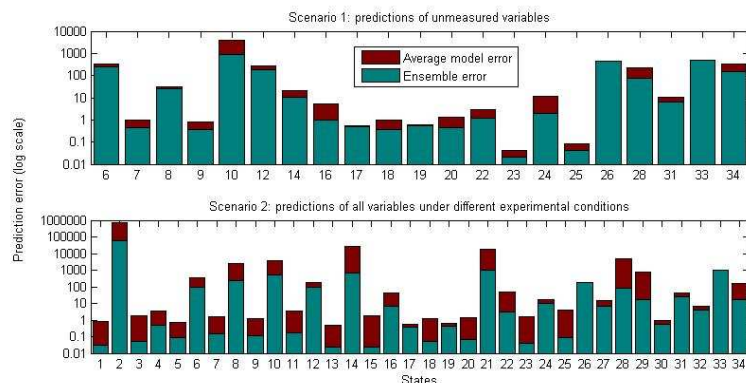


Fig. 8 - Bar plots of the errors made by ensemble prediction for each state (metabolite concentration) vs. the average error of the individual models in the ensemble. The upper panel shows results for the first scenario, where the models were calibrated using data from 13 variables; errors for the remaining 21 unmeasured states are displayed. The lower panel shows results for the second scenario, where a new experimental condition was tested; in this case the errors are displayed for all the 34 states.

Finally, another aspect of the methodology that is open to discussion is whether to include in the ensemble all the available models, or only those calibrated with the meta-parameter approach. The difference between both options can be analyzed according to two criteria: (1) the resulting ensemble prediction error, and (2) how the prediction error for each variable correlates with the ensemble dissensus for that variable. We computed quantities (1) and (2) for ensembles built in two ways: (A) including all the available models, and (B) including only models calibrated with the meta-parameter approach. We found that the differences between both cases were small (data not shown). Furthermore, the ‘best’ strategy was different in the scenarios considered in sections 3.3 (predicting time courses of unmeasured variables) and 3.4 (predicting time courses of all variables in a different experimental condition). To sum up, the consensus approach provided good results for ensembles built with and without the meta-parameter approach, and the differences between both cases are small. Given that all the calibrated models yielded a good fit to the data, we found no reason to leave out of the ensemble the ones that had been calibrated without meta-parameters.

5 Conclusions

We have presented a method for making high-quality predictions in large-scale dynamic models. The method introduces two computational techniques, which we have referred to as the meta-parameter approach and the consensus approach. Its use has been demonstrated on a biotechnological application, a batch process for recombinant

protein production in Chinese Hamster ovary (CHO) cells. After building a metabolic model of CHO cells, we reduced the number of its parameters by applying the meta-parameter approach without reducing the model structure. This decreased the risk of over-fitting and accelerated the parameter estimation procedure, which produced an ensemble of models with different parameterizations. This ensemble was used as a basis for application of the consensus approach, whose core idea is to consider the coincidence among predictions of the models in the ensemble as an indication of the reliability of these predictions. Using a measure of the relative distance between the average prediction of all models and the true value, the procedure quantifies the degree of confidence in the predictions of the time evolution of each of the state variables (which can be e.g. metabolite concentrations, as in the example application). This quantitative indication can be used to decide where it is most efficient to make an additional measurement effort, that is, which state variables should be actually measured and for which we can rely on the model simulations. In summary, it has been shown how the combination of the meta-parameter approach and of the consensus approach can be a valuable tool for improving the identifiability and applicability of complex dynamic models.

Acknowledgements. This work was supported by the EU project “BioPreDyn” (EC FP7-KBBE-2011-5, grant 289434) and by the Spanish MINECO and the European Regional Development Fund (ERDF; project “MultiScales”, DPI2011-28112-C04-03). We further would like to thank Michael Rempel for valuable technical support.

References

1. Kremling, A., Saez-Rodriguez, J.: Systems biology—an engineering perspective. *J. Biotechnol.* 129, 329–351 (2007)
2. Antony, P., Balling, R., Vlassis, N.: From systems biology to systems biomedicine. *Curr. Opin. Biotech.*, 23(4), 604–608, (2012)
3. Medina, M. Á.: Systems biology for molecular life sciences and its impact in biomedicine. *Cell. Mol. Life Sci.*, 70(6), 1035–1053, doi:10.1007/s00018-012-1109-z (2013)
4. De Maria, C., Grassini, D., Vozzi, F., Vinci, B., Landi, A., Ahluwalia, A., Vozzi, G.: HEMET: Mathematical model of biochemical pathways for simulation and prediction of HEpatocyte METabolism. *Comput. Meth. Prog. Biomed.* 92:121–134, doi: 10.1016/j.cmpb.2008.06.004 (2008)
5. Zakharov, M.N., Bhasin, S., Szafran, A.T., Mancini, M.A., Jasuja, R.: Numerical framework to model temporally resolved multi-stage dynamic systems. *Comput. Meth. Prog. Biomed.* 108:750–759, doi: 10.1016/j.cmpb.2012.04.010 (2012)
6. Carinhas, N., Oliveira, R., Alves, P.M., Carrondo, M.J.T., Teixeira, A.P.: Systems biotechnology of animal cells: the road to prediction. *Trends Biotech.* 30(7): 377–385 (2012)
7. Hattersley, J.G., Pérez-Velázquez, J., Chappell, M.J., Bearup, D., Roper, D., Dowson, C., Bugg, T., Evans, N.D.: Indistinguishability and identifiability of kinetic models for the MurC reaction in peptidoglycan biosynthesis. *Comput. Meth. Prog. Biomed.* 104(2):70–80, doi: 10.1016/j.cmpb.2010.07.009 (2011)
8. Villaverde, A.F., Banga, J.R.: Reverse engineering and identification in systems biology: strategies, perspectives and challenges. *J. R. Soc. Interface* 11: 20130505 (2014)

9. Sontag, E.: For differential equations with r parameters, $2r + 1$ experiments are enough for identification. *J. Nonlin. Sci.* 12:553–583 (2002)
10. Kaltenbach, H.M., Dimopoulos, S., Stelling, J.: Systems analysis of cellular networks under uncertainty. *FEBS letters* 583(24), 3923–3930 (2009)
11. Parker, W.S.: Ensemble modeling, uncertainty and robust predictions. *WIREs Clim. Change* 4(3):213–223 (2013)
12. Smith, T., Ross, A., Maire, N., Chitnis, N., Studer, A., Hardy, D., Brooks, A., Penny, M., Tanner, M.: Ensemble Modeling of the Likely Public Health Impact of a Pre-Erythrocytic Malaria Vaccine. *PLoS Medicine* 9(1): e1001157 (2012)
13. Kauffman, S.: A proposal for using the ensemble approach to understand genetic regulatory networks. *J. Theor. Biol.* 230(4), 581–590 (2004)
14. Tran, L.M., Rizk, L.M., Liao, J.C.: Ensemble Modeling of Metabolic Networks. *Biophys. J.* 95(12), 5606–5617 (2008)
15. Kuepfer, L., Peter, M., Sauer, U., Stelling, J.: Ensemble modeling for analysis of cell signaling dynamics. *Nature Biotechnol.* 25, 1001–1006 (2007)
16. Tan, Y., Liao, J.C.: Metabolic ensemble modeling for strain engineers. *Biotechnol. J.* 7, 343–353 (2012)
17. Bever, C.: Selecting high-confidence predictions from ordinary differential equation models of biological networks. PhD Thesis, MIT (2008)
18. Cedersund, G.: Conclusions via unique predictions obtained despite unidentifiability – new definitions and a general method. *FEBS J.* 279(18), 3513–3527 (2012)
19. Villaverde, A. F., Bongard, S., Mauch, K., Müller, D., Balsa-Canto, E., Schmid, J., Banga, J. R. High-Confidence Predictions in Systems Biology Dynamic Models. In 8th International Conference on Practical Applications of Computational Biology & Bioinformatics (PACBB 2014) (pp. 161–171). Springer International Publishing (2014)
20. Chassagnole, C., Noisommit, N., Schmid, J.W., Mauch, K., Reuss, M.: Dynamic modeling of the central carbon metabolism of *E. coli*. *Biotechnol. Bioeng.* 79(1), 53–73 (2002)
21. Visser, D., Heijnen, J.J.: Dynamic simulation and metabolic redesign of a branched pathway using linlog kinetics. *Metab. Eng.* 5(3), 164–176 (2003)
22. Egea, J.A., Martí R., Banga, J.R.: An evolutionary method for complex process optimization. *Comp. Oper. Res.* 37(2), 315–324 (2010)
23. Villaverde, A.F., Ross, J., Morán, F., Balsa-Canto, E., Banga, J.R.: Use of a generalized Fisher equation for global optimization in chemical kinetics. *J. Phys. Chem. A* 115 (30), 8426–8436 (2011)
24. Villaverde, A.F., Egea, J.A., Banga, J.R.: A cooperative strategy for parameter estimation in large scale systems biology models. *BMC Syst. Biol.* 6:75 (2012)
25. Balsa-Canto, E., Banga, J.R.: AMIGO, a toolbox for Advanced Model Identification in systems biology using Global Optimization. *Bioinformatics* 27(16), 2311–2313 (2011)
26. Pinho, J., Sobral, J.L., Rocha, M.: Parallel evolutionary computation in bioinformatics applications. *Comput. Meth. Prog. Biomed.* 110(2):183–191, doi: 10.1016/j.cmpb.2012.10.001 (2013)
27. Ahn, W.S., Antoniewicz, M.R.: Towards dynamic metabolic flux analysis in CHO cell cultures. *Biotechnol. J.* 7(1), 61–74 (2012)
28. Villaverde, A.F., Henriques, D., Smallbone, K., Bongard, S., Schmid, J., Cicin-Sain, D., Crombach, A., Saez-Rodriguez, J., Mauch, K., Balsa-Canto, E., Mendes, P., Jäger, J., Banga, J.R.: BioPreDyn-bench: a suite of benchmark problems for dynamic modelling in systems biology. *BMC Syst. Biol.*, in press (2015)